Volume 183, number 2 FEBS 2489 April 1985

Cloning and sequencing of the *Pseudomonas aeruginosa* PAK pilin gene

Brittan L. Pasloske, B. Brett Finlay and William Paranchych

Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Cananda

Received 22 February 1985

A 1.2-kilobase (kb) *Hin*dIII restriction fragment containing the pilin gene from *Pseudomonas aeruginosa* PAK has been cloned and sequenced. The pilin protein is 144 amino acids in length with a positively charged leader sequence of 6 amino acids. There is probably only one copy of the gene per chromosome.

Pseudomonas aeruginosa *pilin gene* *Chromosome location* *DNA sequence* *Amino acid sequence* *Leader sequence* *Leader cleavage*

1. INTRODUCTION

Pseudomonas aeruginosa is an opportunistic pathogen which often invades hosts who have been weakened by diseases such as leukemia and cystic fibrosis or by tissue damage caused by severe burns. Recent studies have indicated that the adherence of *P. aeruginosa* to mucosal and epithelial surfaces is mediated by filamentous adherence structures known as pili and that this adherence phenotype is necessary for bacterial colonization [1].

Many strains of *P. aeruginosa* produce chromosomally encoded pili which are polar, flexible filaments of 5.2 nm diameter and an average of 2.5 μm in length [2,3]. The pili are retractable and promote infection by various pilus-specific bacteriophages [4]. A primitive form of movement known as twitching motility [5] as well as the adhesion of *Pseudomonas* bacteria to mammalian tissues [1,6] have been attributed to pili.

Here we present the complete nucleotide sequence of a 1.2 kilobase *Hin*dIII restriction fragment from *P. aeruginosa* PAK encoding the entire pilin gene. The translated nucleotide sequence is compared to the published amino acid sequence for *P. aeruginosa* PAK pilin. We also present preliminary evidence that the pilin gene exists as a single copy within the chromosome.

2. EXPERIMENTAL

Oligonucleotide probes of CA(A/G)TA(T/C)-CA(A/G)AA(T/C)TA(T/C)GT were purchased from New England Biolabs (MA) and end-labelled with T_4 polynucleotide kinase and [γ-^{32}P]ATP [7]. Southern blots [8] were prehybridized with 0.9 M NaCl, 0.09 M Tris (pH 7.5), 0.006 M EDTA, 0.5% SDS, 0.05 μg/ml sheared calf thymus DNA and 5 × Denhardt's solution [0.02% Ficoll, 0.02% bovine serum albumin, 0.02% polyvinylpolypyrrolidone in 1 × SSC (0.15 M NaCl, 0.015 M sodium citrate, pH 7)] at 34°C for 4 h, then hybridized under the same conditions with labelled synthetic oligonucleotides for 8 h. Blots were washed with 6 × SSC, 0.1% SDS at room temperature for 1 h and exposed to X-ray film. Double-stranded DNA was nick-translated and used to probe Southern blots as described in [9].

P. aeruginosa PAK chromosome was prepared using the procedure according to Coleman et al. [10]. Complete *Hin*dIII restriction digests of *P. aeruginosa* PAK DNA were ligated into pUC 8 and transformed into the *E. coli* strain JM83 [11]. Colonies positive for inserts were plated onto nitrocellulose, grown, and lysed according to [9]. The nitrocellulose filters were washed free of cell debris [7] and then probed with the labelled synthetic oligonucleotides as above.

Published by Elsevier Science Publishers B.V. (Biomedical Division)

Recombinant plasmids were isolated in large quantities using the cleared lysate method described by Crosa and Falkow [12], followed by CsCl density centrifugation while smaller amounts were obtained by the method of Birnboim and Doly [13].

Nucleotide sequence analysis was performed using the Sanger dideoxy method [14] using M13 mp18/19 [15] or pUC 8 [16].

3. RESULTS AND DISCUSSION

Several methods were utilized unsuccessfully in attempting to clone the *P. aeruginosa* PAK pilin gene. Whole cell immunoblot assays that were successful with the *Neisseria gonorrhoeae* pilin gene [17] did not work for *P. aeruginosa*. It is known that the first 30 amino acids at the amino terminus of *P. aeruginosa* PAK pili and *N. gonorrhoeae* pili are homologous (differing by 7 amino acids) [18]. On this basis, it was predicted that the DNA encoding the *N. gonorrhoeae* pilin protein might share enough homology to permit its use as a probe for the *P. aeruginosa* pilin gene. A chimera containing the *N. gonorrhoeae* pilin gene (kindly provided by Dr M. Koomey, Rockefeller University) was tested in this capacity. However, this probe failed to identify any homologous fragments with a Southern blot of *P. aeruginosa* chromosome, although it did identify 4 bands from a chromosomal *Eco*RI digest of a clinical isolate of *N. gonorrhoeae* (not shown).

Using the known *Pseudomonas aeruginosa* PAK pilin sequence [19], a synthetic oligonucleotide was identified as a potential DNA probe for the pilin gene. The nucleotide sequence corresponding to residues 23–28 was chosen since it represented an area of low codon ambiguity. Several restriction digests of *P. aeruginosa* PAK chromosome were performed and a Southern blot of these digests made. When this blot was probed with the synthetic oligonucleotide pool and washed under low stringency conditions (see section 2), a single, distinct band from each digest was found to contain homology in all cases. Since a *Hin*dIII fragment of approx. 1.2 kb showed homology with the synthetic oligonucleotides, and because the *Hin*dIII digest contained few fragments in this size range (fig.1), a complete *Hin*dIII chromosomal digest was used for ligation into the vector pUC 8.

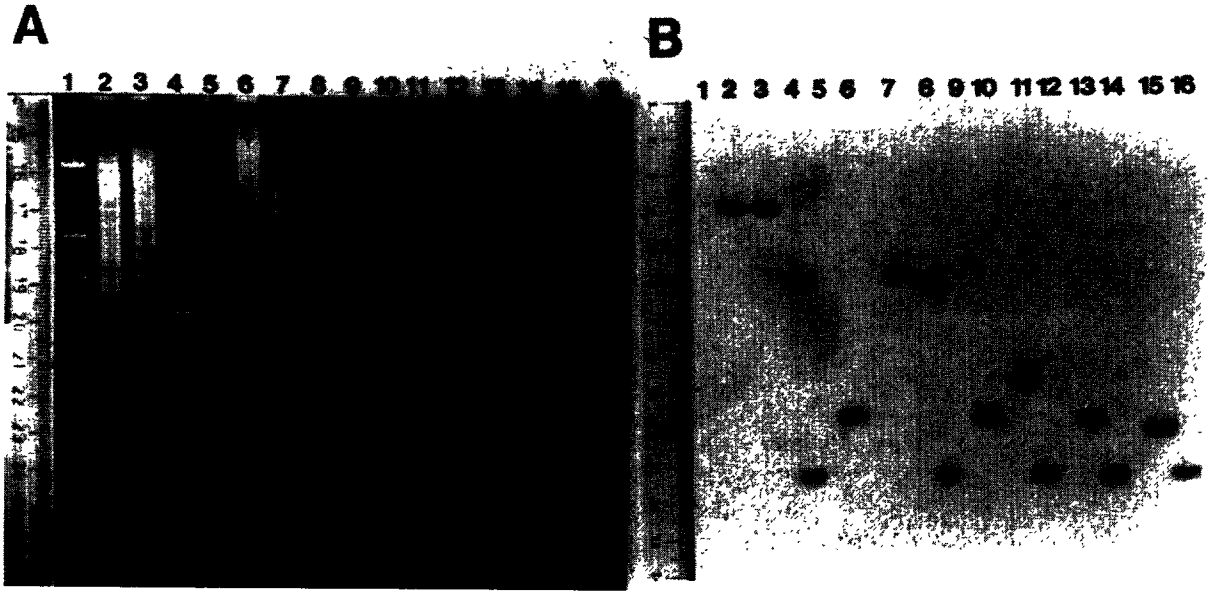

Fig.1. (A) 0.6% agarose gel of various chromosomal digests of *P. aeruginosa* PAK. (B) Autoradiograph of Southern blot of (A) hybridized with 610 bp *Pst*I fragment containing the first 102 N-terminal amino acids of the pilin gene. Lanes: (1) λ/*Hin*dIII, (2) *Eco*RI, (3) *Bam*HI, (4) *Sal*I, (5) *Pst*I, (6) *Hin*dIII, (7) *Eco*RI/*Bam*HI, (8) *Eco*RI/*Sal*I, (9) *Eco*RI/*Pst*I, (10) *Eco*RI/*HindIII, (11) Bam*HI/*Sal*I, (12) *Bam*HI/*Pst*I, (13) *Bam*HI/*Hin*dIII, (14) *Sal*I/*Pst*I, (15) *Sal*I/*Hin*dIII, (16) *Pst*I/*Hin*dIII.

Chimeras containing inserts were tested by probing colony blots with the synthetic oligonucleotide pool. Five out of 250 colonies showed homology with this probe pool. DNA from one positive recombinant, pBP001, was isolated using a rapid plasmid procedure [13], digested with *Hin*dIII, and shown to contain a 1.2 kb fragment which had been inserted into pUC 8. This DNA fragment was also sequenced using the oligonucleotide pool as a primer [14,16] and shown to contain the N-terminal region of *P. aeruginosa* PAK pilin.

The entire 1224 base pair *Hin*dIII fragment was sequenced using the dideoxy method with M13 mp19 and pUC 8 clones (see fig.2 for sequencing strategy, fig.3 for sequence). The complete *P. aeruginosa* PAK pilin gene was found within this *Hin*dIII fragment (nucleotides 457–906) and was the same as the published protein sequence [19], except that it contained a 6 amino acid leader sequence and a carboxy-terminus arginine, instead

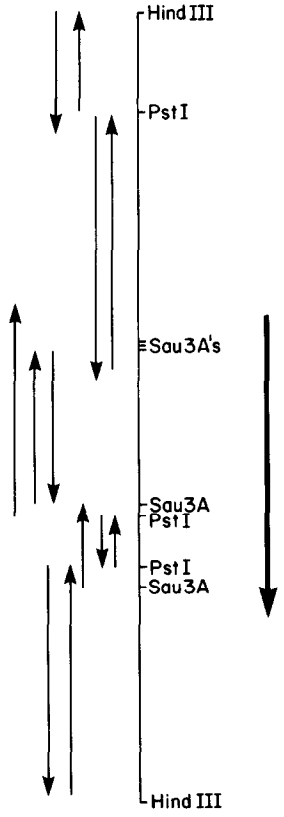

Fig.2. Sequencing strategy for the 1.2 kb *Hin*dIII fragment (pBP001) containing the pilin gene. Bold arrow indicates location of pilin gene and direction of transcription. The direction of the sequenced DNA fragments is shown by arrows beneath the restriction map.

```
                                   30                  60
AAGCTTTCCTGTCCAGGTGTTCAGGTCGCAATAGGCGATGCCGAACTGCTCGGCGGACAA
HindIII
                                   90                  120
TTCGGCCAGGGCCAGGCCGCTTACCAGCTTGTTCTGCACCAGGTGCGTCACCAGCGACAG

                                  150                  180
CTTGTTGCGCTGCGCCTGCGTCTGGGCCTGCAGGGCGGTCTTTTCGTCGAGTAGATTGGC
                                 PstI
                                  210                  240
TTGGACGAGCTGTCGGGACAGACCGCTCAGTTGGATGCTGTCGTTCATGGAGAAGGAAAT

                                  270                  300
CGCAGAGGGCTATTGAAGTGCCTTATAACGCAGATAACAGGGTCTGCCAAATCGGGGGAG

                                  330                  360
TCCGGCTGTCAAAAAGTGTCACATCCTGTCCGTTTTAAGTTTGAGTCTCCTCGGGAAGGG

                                  390                  420
CATAAGACGGCCTTGCTCGCTATGTGAACGGTATTTGGCATGGTAAGTGCTTGGTAGGGT

                                  450                  480
TAGGCGTTAGGCCTATACATATCAATGGAGATATTCATGAAAGCTCAAAAAGGCTTTACC
                                       MetLysAlaGlnLysGlyPheThr
                                  510                  540
TTGATCGAACTGATGATCGTGGTTGCGATCATCGGTATCTTGGCTGCAATTGCCATTCCT
LeuIleGluLeuMetIleValValAlaIleIleGlyIleLeuAlaAlaIleAlaIlePro
  Sau3A        Sau3A        Sau3A
                                  570                  600
CAGTATCAGAATTATGTAGCTCGTTCGGAAGGCGCATCTGCTCTTGCTTCGGTCAATCCG
GlnTyrGlnAsnTyrValAlaArgSerGluGlyAlaSerAlaLeuAlaSerValAsnPro

                                  630                  660
TTGAAGACTACCGTTGAAGAGGCGCTTTCTCGTGGTTGGAGCGTGAAGAGCGGTACAGGT
LeuLysThrThrValGluGluAlaLeuSerArgGlyTrpSerValLysSerGlyThrGly

                                  690                  720
ACAGAGGACGCTACTAAGAAAGAGGTTCCTCTGGGGGTGGCGGCAGATGCTAACAAACTG
ThrGluAspAlaThrLysLysGluValProLeuGlyValAlaAlaAspAlaAsnLysLeu

                                  750                  780
GGTACTATCGCACTCAAACCCGATCCTGCTGATGGTACTGCAGATATCACTTTGACTTTC
GlyThrIleAlaLeuLysProAspProAlaAspGlyThrAlaAspIleThrLeuThrPhe
                    Sau3A            PstI
                                  810                  840
ACTATGGGCGGTGCAGGACCGAAGAATAAAGGGAAAATTATTACCCTGACTCGTACTGCA
ThrMetGlyGlyAlaGlyProLysAsnLysGlyLysIleIleThrLeuThrArgThrAla
                                                      PstI
                                  870                  900
GCTGATGGTCTCTGGAAGTGCACCAGTGATCAGGATGAGCAGTTTATTCCGAAAGGTTGC
AlaAspGlyLeuTrpLysCysThrSerAspGlnAspGluGlnPheIleProLysGlyCys
                                 Sau3A
                                  930                  960
TCTAGGTAATTTTGTAAGCGCTGGATGGTGTTGAACTATCGATGATTTGATGATAGTTTC
SerArgEnd

                                  990                 1020
CACGCCGCCGCCGGATTAGCTCAGTCGGTAGAGCAGCTCATTCGTAATGAGAAGGTCGGG

                                 1050                 1080
GGTTCGATTCCTCTATCCGGCACCAGTCGCATAAAAAGCCCCGCTTCGGCGGGGCTTTTT

                                 1110                 1140
TTCGCCTGCGATTCGTTCAGAGGGTGAGGCGCATGGACAGGTCGACGGCCCTGACGTCCT

                                 1170                 1200
TGGTCAGGGTGCCGATGGAGATGTAGTCGACGCCGGTCTCGGCGATGTTGCGCAGGGTGC

                                 1230
TTTCGTTGATCCCGCCGGAAGCTT
       Sau3A      HindIII
```

Fig.3. Nucleotide sequence of the 1224 base pair *Hin*dIII fragment (pBP001). The mature *P. aeruginosa* PAK pilin protein sequence with leader sequence is shown with the putative cleavage site of the leader sequence represented by the arrow and the proposed Shine-Dalgarno sequence underlined.

of lysine. This C-terminus variation is not an error in the protein sequence, but represents a real difference between pilin derived from a multi-piliated mutant of *P. aeruginosa* PAK (PAK/2Pfs), which was used for protein sequencing studies [19], and the wild type strain of *P. aeruginosa* PAK, from which the present pilin gene was cloned.

According to the nucleotide sequence, the mature pilin molecule has 6 amino acids removed from its amino terminus. Most bacterial proteins which are to be secreted or inserted into the outer membrane have an amino-terminal extension which is subsequently removed [21]. This signal sequence is usually characterized by a hydrophobic core of 12–20 amino acids with one or more positively charged amino acids preceding this region. The *P. aeruginosa* PAK pilin leader sequence lacks this hydrophobic core but the N-terminal region of the mature pilin is hydrophobic and could substitute for the hydrophobic region missing in the signal sequence. Four nucleotides (GAGG) 6 positions upstream of the ATG start codon of pilin match the consensus sequence for a ribosome binding site [22]. The proposed translational start site for the PAK pilin gene is shown at nucleotide position 457. This was chosen in preference to position 445 (the only other possible start codon) because it is preceded by a potential ribosome binding site, while such a ribosome binding site is not evident in the region preceding position 445. The 1224 bp *Hin*dIII fragment also contains an open reading frame of 111 amino acids extending from the *Hin*dIII site (position 1) to nucleotide 335. Perpaps this is a C-terminal portion of a protein found upstream of the pilin gene on the chromosome. No other reading frames of any significance were found in either direction on this fragment.

Other infectious organisms such as *Bacteroides nodosus* and *N. gonorrhoeae* have pilin proteins which have a homologous amino-terminal region [23,24]. The nucleotide sequences for these pilin genes have also been determined [23,24] and they have similar short, positively charged leader sequences with the last 3 residues of the leader conserved among all 3 leader sequences. The *P. aeruginosa* PAK leader sequence has two positive charges within 6 residues compared to one positive charge found within 7 residues in *N. gonorrhoeae* and *B. nodosus*. These charges may be important in preventing aggregation of pilin subunits before pilus assembly occurs.

Probing the same Southern blot as used for the synthetic oligonucleotides with the 610 bp *Pst*I fragment of the cloned DNA (containing the signal sequence and the first 96 amino acids) resulted in a blot identical to the one probed with the oligonucleotide pool (fig.1). With all digests only a single band appeared. These data indicate that the pilin gene is found as a single copy in the genome, or within a very large repeated segment. This is in contrast with *N. gonorrhoeae*, which has at least 3 copies in its genome [24].

When the synthetic oligonucleotide pool was hybridized to various digests of *P. aeruginosa* PAO chromosomal DNA, no homology was detected, even though these proteins have similar amino acid sequences in the amino terminal region (Sastry, P. et al., unpublished). It appears that the PAO pilin gene contains differences in amino acid sequence in the region of residues 23–28 which prevent the oligonucleotide pool from hybridizing. However, when the *Pst*I 610 bp fragment from PAK was used as a probe with PAO, a single band with detectable homology was found in all digests. We have cloned the *P. aeruginosa* PAO pilin gene and are currently determining its sequence.

ACKNOWLEDGEMENTS

This work was supported by the Medical Research Council of Canada. B.L.P. and B.B.F. are recipients of a Studentship from the Alberta Heritage Foundation for Medical Research.

REFERENCES

[1] Woods, D., Strauss, D., Johanson, W., Berry, V. and Bass, J. (1980) Infect. Immun. 29, 1146–1151.
[2] Folkhard, W., Marvin, D., Watts, T. and Paranchych, W. (1981) J. Mol. Biol. 149, 79–93.
[3] Bradley, D. (1972) Genet. Res. 19, 39–51.
[4] Bradley, D. and Pitt, T. (1974) J. Gen. Virol. 24, 1–15.
[5] Bradley, D. (1980) Can. J. Microbiol. 26, 146–154.
[6] Reichert, R., Das, N. and Zam, Z. (1983) Curr. Eye Res. 2, 289–293.
[7] Woods, D. (1984) in: Focus, pp.1–2, vol.6, Bethesda Research Laboratories, Gaithersburg, MD.

[8] Southern, E. (1975) J. Mol. Biol. 98, 503–517.
[9] Maniatis, T., Fritsch, E. and Sambrook, J. (1982) in: Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[10] Coleman, K., Dougan, G. and Arbuthnot, J.P. (1983) J. Bacteriol. 153, 909–915.
[11] Viera, J. and Messing, L. (1982) Gene 19, 259–268.
[12] Crosa, J. and Falkow, S. (1981) in: Plasmids (Gerhardt, R. et al. eds) Manual of Methods for General Bacteriology, American Society for Microbiology, pp.266–282, Washington, DC.
[13] Birnboim, H. and Doly, J. (1979) Nucleic Acids Res. 7, 1513–1523.
[14] Sanger, F., Nicklen, S. and Coulson, A. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[15] Messing, J., Crea, R. and Seeburg, P. (1981) Nucleic Acids Res. 9, 309–322.
[16] Wallace, R., Johnson, M., Suggs, S., Miyoshi, K., Bhatt, R. and Itakura, K. (1981) Gene 16, 21–26.
[17] Meyer, J., Mlawer, N. and So, M. (1982) Cell 30, 45–52.
[18] McKern, N., O'Donnell, I., Inglis, A., Stewart, D. and Clark, B. (1983) FEBS Lett. 164, 149–153.
[19] Sastry, P., Pearlstone, J., Smillie, L. and Paranchych, W. (1983) FEBS Lett. 151, 253–256.
[20] Bradley, B. (1974) Virology 58, 149–163.
[21] Pages, J. (1983) Biochimie 65, 531–541.
[22] Shine, J. and Dalgarno, L. (1975) Nature 254, 34–38.
[23] Elleman, T. and Hoyne, P. (1984) J. Bacteriol. 160, 1184–1187.
[24] Meyer, T., Billyard, E., Haas, R., Storzbach, S. and So, M. (1984) Proc. Natl. Acad. Sci. USA 81, 6110–6114.